AMENDMENTS

In the Claims

The following is a marked-up version of the claims with the language that is underlined ("___") being added and the language that contains strikethrough ("~~___~~") being deleted:

1. (Currently Amended) A method comprising:

receiving a first email message from a simple mail transfer protocol ~~(SMTP)~~ server, the first email message comprising displaying characters and non-displaying characters, the non-displaying characters including non-displaying comments and non-displaying control characters; the first email message further comprising:

a 32-bit string indicative of a length of the first email message;

a text body;

~~an SMTP~~ a simple mail transfer protocol email address that includes a user name and a domain name;

an attachment;

searching for the non-displaying characters in the first email message;

removing the non-displaying characters, including the non-displaying comments and the non-displaying control characters;

determining non-alphabetic displaying characters in the first email message, where determining the non-alphabetic displaying characters includes a per-character analysis that recursively determines for each character whether:

a character is a non-alphabetic character;

if the character is a non-alphabetic character, whether the character is a space;

if the character is a space, determine whether the space is adjacent to a solitary "i" or "a"; ~~and~~

in response to a determination that the space is not adjacent to a solitary “i" or “a”, deleting the non-alphabetic character; and

if the non-alphabetic character is not a space, filtering the determined non-alphabetic displaying characters from the first email message;

generating a phonetic equivalent for each word that includes only alphabetic displaying characters that has a phonetic equivalent;

tokenizing the phonetic equivalents in a displaying portion of the text body to generate a plurality of body tokens representative of words in the text body;

tokenizing the ~~SMTP~~ simple mail transfer protocol email address to generate an address token representative of the ~~SMTP~~ simple mail transfer protocol email address;

tokenizing the domain name to generate a domain token that is representative domain name;

tokenizing the attachment to generate an attachment token that is representative of the attachment, wherein tokenizing comprises:

generating a 128-bit MD5 hash of the attachment;

appending the 32-bit string to the generated MD5 hash to produce a 160-bit number; and

UUencoding the 160-bit number to generate the attachment token representative of the attachment;

determining a corresponding spam probability value for each of the plurality of body tokens, the address token, the domain token, and the attachment token;

determining whether at least one of the plurality of body tokens, the address token, the domain token, and the attachment token is present in a database of tokens and, in response to a determination that at least one of the plurality of body tokens, the address token, the domain token, and the attachment token is present in the database of tokens:

updating the spam probability value of the plurality of body tokens, the address token, the domain token, and the attachment token; and

sorting the plurality of body tokens, the address token, the domain token, and the attachment token in accordance with the corresponding spam probability value to determine a predefined number of interesting tokens, the predefined number of interesting tokens being a subset of the plurality of body tokens, the address token, the domain token, and the attachment token;

classifying the plurality of body tokens, the address token, the domain token, and the attachment token as spam, non-spam, or neutral;

selecting the predefined number of interesting tokens, to create selected interesting tokens, the selected interesting tokens being the plurality of body tokens, the address token, the domain token, and the attachment token having a greatest non-neutral probability values;

performing a Bayesian analysis on the selected interesting tokens to generate a spam probability;

categorizing the first email message as a function of the spam probability; and

filtering a second email message.

2. – 5. (Canceled)

6. (Currently Amended) A method comprising:

receiving, at a computing device, a first email message comprising a text body, ~~an SMTP~~ a simple mail transfer protocol email address, an attachment, and a domain name corresponding to the ~~SMTP~~ simple mail transfer protocol email address, the text body including displaying characters and non-displaying characters;

searching for the non-displaying characters in the first email message;

removing the searched non-displaying characters, including non-displaying comments and non-displaying control characters;

determining non-alphabetic displaying characters in the first email message, where determining the non-alphabetic displaying characters includes a per-character analysis that recursively determines for each character whether:

a character is a non-alphabetic character;

if the character is a non-alphabetic character, whether the character is a space;

if the character is a space, determine whether the space is adjacent to a solitary "i" or "a";

in response to a determination that the space is not adjacent to a solitary "i" or "a", deleting the non-alphabetic character; and

if the non-alphabetic character is not a space, filtering the determined non-alphabetic displaying characters from the first email message;

tokenizing the ~~SMTP~~ simple mail transfer protocol email address to generate an address token representative of the displaying characters of the ~~SMTP~~ simple mail transfer protocol email address;

tokenizing the attachment to generate an attachment token that is representative of the attachment;

tokenizing the domain name to generate a domain token representative of the domain name;

determining a corresponding spam probability value from the address token, the attachment token, and the domain token;

determining whether at least one of the address token, the attachment token, and the domain token is present in a database of tokens and, in response to a determination that at

least one of the address token, the attachment token, and the domain token is present in the database of tokens:

updating the spam probability value of at least one of the address token, the attachment token, and the domain token;

sorting the address token, the attachment token, and the domain token in accordance with the corresponding spam probability value to determine a predefined number of interesting tokens, the predefined number of interesting tokens being a subset of the address token, the attachment token, and the domain token; and

filtering a second email message.

7. – 10. (Canceled)

11. (Currently Amended) The method of claim 6, wherein determining the spam probability comprises:

assigning an address spam probability value to the address token representative of the ~~SMTP~~ simple mail transfer protocol email address;

assigning a domain spam probability value to the domain token representative of the domain name; and

generating a Bayesian probability value using the address spam probability and the domain spam probability assigned to the address token and the domain token.

12. (Previously Presented) The method of claim 11, wherein determining the spam probability further comprises:

comparing the Bayesian probability value with a predefined threshold value.

13. (Previously Presented) The method of claim 12, wherein determining the spam probability further comprises:

categorizing the first email message as spam in response to the Bayesian probability value being greater than the predefined threshold.

14. (Previously Presented) The method of claim 12, wherein determining the spam probability further comprises:

categorizing the first email message as non-spam in response to the Bayesian probability value being not greater than the predefined threshold.

15. (Canceled)

16. (Previously Presented) The method claim 6, wherein receiving the first email message further comprises:

receiving the first email message including a text body.

17. (Previously Presented) The method of claim 16, further comprising:

tokenizing the words in the text body to generate body tokens representative of the words in the text body.

18. (Canceled)

19. (Previously Presented) The method of claim 17, wherein determining the spam probability comprises:

assigning a body spam probability value to each of the body tokens representative of the

words in the text body;

assigning an attachment spam probability value to the attachment token representative of the attachment; and

generating a Bayesian probability value using the body spam probability value and the attachment spam probability value assigned to the body tokens and the attachment token.

20. (Previously Presented) The method of claim 19, wherein determining the spam probability further comprises:

comparing the Bayesian probability value with a predefined threshold value.

21. (Previously Presented) The method of claim 20, wherein determining the spam probability further comprises:

categorizing the first email message as spam in response to the Bayesian probability value being greater than the predefined threshold.

22. (Previously Presented) The method of claim 20, wherein determining the spam probability further comprises:

categorizing the first email message as non-spam in response to the Bayesian probability value being not greater than the predefined threshold.

23. (Currently Amended) A system comprising:

a memory component that stores at least the following:

email receive logic configured to receive a first email message comprising ~~an SMTP~~ a simple mail transfer protocol email address, a domain name corresponding to the ~~SMTP~~ simple mail transfer protocol email address, and an attachment, the first email message

further including displaying characters and non-displaying characters;

searching logic configured to search for the non-displaying characters in the first email message;

removing logic configured to remove the non-displaying characters, including non-displaying comments and non-displaying control characters;

first determine logic configured to determine non-alphabetic displaying characters in the first email message, where determining the non-alphabetic displaying characters includes a per-character analysis that recursively determines for each character whether:

a character is a non-alphabetic character;

if the character is a non-alphabetic character, whether the character is a space;

if the character is a space, determine whether the space is adjacent to a solitary “i” or “a”;

in response to a determination that the space is not adjacent to a solitary “i" or “a”, deleting the non-alphabetic character; and

if the non-alphabetic character is not a space, filtering the determined non-alphabetic displaying characters from the first email message;

tokenize logic configured to tokenize the ~~SMTP~~ simple mail transfer protocol email address to generate an address token representative of the ~~SMTP~~ simple mail transfer protocol email address;

tokenize logic configured to tokenize the attachment to generate an attachment token that is representative of the attachment;

tokenize logic configured to tokenize the domain name to generate a domain token representative of the domain name;

analysis logic configured to determine a corresponding spam probability value

from the address token, the attachment token, and the domain token; and

second determine logic configured to determine whether at least one of the address token, the attachment token, and the domain token is present in a database of tokens and, in response to a determination that at least one of the address token, the attachment token, and the domain token is present in the database of tokens:

update the corresponding spam probability value of the address token, the attachment token, and the domain token;

sort the address token, the attachment token, and the domain token in accordance with the corresponding spam probability value to determine a predefined number of interesting tokens, the predefined number of interesting tokens being a subset of the address token, the attachment token, and the domain token, wherein only displaying characters are tokenized; and

filter a second email message.

24. (Canceled)

25. (Currently Amended) A non-transitory computer-readable storage medium that includes a program that, when executed by a computer, performs at least the following:

receive a first email message comprising ~~an SMTP~~ a simple mail transfer protocol email address, a domain name corresponding to the ~~SMTP~~ simple mail transfer protocol email address, and an attachment, the first email message further including displaying characters and non-displaying characters;

search for non-displaying characters in the first email message;

remove the non-displaying characters, including non-displaying comments and non-displaying control characters;

determine non-alphabetic displaying characters in the first email message, where determining the non-alphabetic displaying characters includes a per-character analysis that recursively determines for each character whether:

a character is a non-alphabetic character;

if the character is a non-alphabetic character, whether the character is a space;

if the character is a space, determine whether the space is adjacent to a solitary "i" or "a";

in response to a determination that the space is not adjacent to a solitary "i" or "a", deleting the non-alphabetic character; and

if the non-alphabetic character is not a space, filtering the determined non-alphabetic displaying characters from the first email message;

tokenize the ~~SMTP~~ simple mail transfer protocol email address to generate an address token representative of the ~~SMTP~~ simple mail transfer protocol email address;

tokenize the attachment to generate an attachment token that is representative of the attachment;

tokenize the domain name to generate a domain token representative of the domain name;

determine a corresponding spam probability value from the address token, the attachment token, and the domain token; and

determine whether at least one of the address token, the attachment token, and the domain token is present in a database of tokens and, in response to a determination that at least one of the address token, the attachment token, and the domain token is present in the database of tokens:

update the corresponding spam probability value of the address token, the attachment token, and the domain token;

sort the address token, the attachment token, and the domain token in accordance with the corresponding spam probability value to determine a predefined number of interesting tokens, the predefined number of interesting tokens being a subset of the generated tokens, wherein only displaying characters are tokenized; and

filter a second email message.

26. (Currently Amended) The non-transitory computer-readable storage medium of claim 25, the program further causing the computer to perform at least the following:

assign an address spam probability value to the address token representative of the ~~SMTP~~ simple mail transfer protocol email address;

assign a domain spam probability value to the domain token representative of the domain name; and

generate a Bayesian probability value using the address spam probability value and the domain spam probability value assigned to the tokens.

27. (Currently Amended) The non-transitory computer-readable storage medium of claim 26, the program further causing the computer to perform at least the following:

compare the Bayesian probability value with a predefined threshold value.

28. (Currently Amended) The non-transitory computer-readable storage medium of claim 27, the program further causing the computer to perform at least the following:

categorize the first email message as spam in response to the Bayesian probability value being greater than the predefined threshold.

29. (Currently Amended) The non-transitory computer-readable storage medium of claim 27, the program further causing the computer to perform at least the following:

categorize the first email message as non-spam in response to the Bayesian probability value being not greater than the predefined threshold.

30. (Currently Amended) A system comprising:

a memory component that stores at least the following:

email receive logic configured to receive a first email message comprising an attachment and an address, the email message further including displaying characters and non-displaying characters;

search logic configured to search for the non-displaying characters in the first email message;

remove logic configured to remove the non-displaying characters, including non-displaying comments and non-displaying control characters;

determine logic configured to determine non-alphabetic displaying characters in the first email message, where determining the non-alphabetic displaying characters includes a per-character analysis that recursively determines for each character whether:

a character is a non-alphabetic character;

if the character is a non-alphabetic character, whether the character is a space;

if the character is a space, determine whether the space is adjacent to a solitary “i” or “a”;

in response to a determination that the space is not adjacent to a solitary “i" or “a”, deleting the non-alphabetic character; and

if the non-alphabetic character is not a space, filtering the determined

non-alphabetic displaying characters from the first email message;

tokenize logic configured to generate at least one attachment token representative of the attachment;

analysis logic configured to determine a corresponding spam probability value from the at least one attachment token; and

database determining logic configured to determine whether the at least one attachment token is present in a database of tokens and, in response to a determination that the at least one attachment token is present in the database of tokens:

update the corresponding spam probability value of the at least one attachment token;

sort the at least one attachment token in accordance with the corresponding spam probability value to determine a predefined number of interesting tokens, the predefined number of interesting tokens being a subset of the at least one attachment token, wherein only displaying characters are tokenized; and

filter a second email message.

31. (Canceled)

32. (Currently Amended) A non-transitory computer-readable storage medium that includes a program that, when executed by a computer, performs at least the following:

receive a first email message comprising an attachment and an address, the first email message further including displaying characters and non-displaying characters;

search for the non-displaying characters in the first email message;

remove the non-displaying characters, including non-displaying comments and non-displaying control characters;

determine non-alphabetic displaying characters in the first email message, where determining the non-alphabetic displaying characters includes a per-character analysis that recursively determines for each character whether:

a character is a non-alphabetic character;

if the character is a non-alphabetic character, whether the character is a space;

if the character is a space, determine whether the space is adjacent to a solitary "i" or "a";

in response to a determination that the space is not adjacent to a solitary "i" or "a", deleting the non-alphabetic character; and

if the non-alphabetic character is not a space, filtering the determined non-alphabetic displaying characters from the first email message;

generate at least one attachment token representative of the attachment;

determine a spam probability value from the at least one attachment token; and

determine whether the at least one attachment token is present in a database of tokens and, in response to a determination that the at least one attachment token is present in the database of tokens:

update the spam probability value of the at least one attachment token;

sort the at least one attachment token in accordance with the spam probability value to determine a predefined number of interesting tokens, the predefined number of interesting tokens being a subset of the generated tokens, wherein only displaying characters are tokenized; and

filter a second email message.

33. (Currently Amended) The non-transitory computer-readable storage medium of claim 32, the program further causing the computer to perform at least the following:

receive the first email message having a text body.

34. (Currently Amended) The non-transitory computer-readable storage medium of claim 33, the program further causing the computer to perform at least the following:

tokenize words in the text body to generate body tokens representative of the words in the text body.

35. (Currently Amended) The non-transitory computer-readable storage medium of claim 34,

assign a body spam probability value to each of the body tokens representative of the words in the text body;

assign an attachment spam probability value to the token representative of the attachment; and

generate a Bayesian probability value using the ~~the~~ attachment spam probability and the body spam probability assigned to the the body tokens and the attachment token.

36. (Currently Amended) The non-transitory computer-readable storage medium of claim 35, the program further causing the computer to perform at least the following:

compare the Bayesian probability value with a predefined threshold value.

37. (Currently Amended) The non-transitory computer-readable storage medium of claim 36, the program further causing the computer to perform at least the following:

categorize the first email message as spam in response to the Bayesian probability value being greater than the predefined threshold.

38. (Currently Amended) The non-transitory computer-readable storage medium of claim 36, the program further causing the computer to perform at least the following:

categorize the first email message as non-spam in response to the Bayesian probability value being not greater than the predefined threshold.

39. (Previously Presented) The method of claim 1, wherein the first email message is received at a computing device.

40. (Canceled)